A method to build unique calibration models for batches of chemical sensor arrays

J. Fonollosa[1,2], R. Rodríguez-Pérez[1], A. Solórzano[1,2], M. Padilla[1], T. Graunke[3], S. Marco[1,2]
[1]Institute for Bioengineering of Catalonia, Baldiri i Reixac 4-8, 08028 Barcelona.
[2]Department of Electronics, Universitat de Barcelona, Martí i Franqués 1, 08028 Barcelona.
[3]ams Sensor Solutions Germany, Gerhard-Kindler-Str. 8, 72770 Reutlingen.
jfonollosa@ibecbarcelona.eu

Summary

Calibration costs are significant for chemical sensor arrays and calibration for replicas of a sensor array needs to be performed individually due to sensor variability, thereby limiting mass-production. To reduce calibration costs, we present a methodology that rejects sensor variability and provides unique calibration models suitable for replicas of a sensor array avoiding individual calibration. We validated our method using a dataset of three different metal oxide gas sensors operating at eight hot plate temperatures. Five copies of said sensor were exposed to six calibration gases at several relative humidity levels. Results showed that it is possible to build unique classification models that keep classification accuracy when applied to new replicas of a sensor array, without the need of calibration transfer strategies.

Keywords: Gas sensor array, calibration, MOX gas sensor, data processing

Motivation

Since the early 90's it is well-known that chemical sensor arrays coupled with multivariate data processing improve the selectivity of individual sensors. However, calibration processes for sensor arrays require larger number of calibration conditions to incorporate cross-sensitivities to the model, and calibration for each sensor array is still required due to sensor variability. As a result, high calibration costs prevent widespread deployment. In contrast to previous approaches that presented calibration transfer strategies, which still require the acquisition of transfer samples [1], we present a methodology that rejects sensor variability and therefore can be extended to replicas of sensing units.

Experimental

The sensor arrays were placed (sequentially) in a 287-ml measurement chamber while gas conditions were changed and sensor conductances acquired. Each sensor array was composed of a different sensor type (AS-MLK, AS-MLC, AS-MLX, all provided by ams). Sensors within each type were operated at eight different temperatures (from 245ºC to 381ºC). The arrays were exposed to carbon monoxide, acetaldehyde, methane, ethanol, nitrogen dioxide, and propane; at different concentration levels and three humidity levels (20, 40, and 60%). Moreover, in order to include sensor drift in the dataset, CO and NO_2 measurements were performed twice, measuring 50 days apart.

Methodology

We tested our methodology in the task of CO identification by means of training Partial Least Squares - Discriminant Analysis (PLS-DA) models. We grouped sensors in arrays of 24 elements: for each sensor type (3) we selected an element at each temperature (8). Therefore, we could build five 24-element ensembles. The model was trained considering four of the arrays and it was tested with the remaining array.

In order to ascertain whether a subset of sensors working at given temperatures leads to higher performance than considering all 24 conditions, we applied feature selection using Genetic Algorithms (GA) [2]. For fitness evaluation of each individual, we built a PLS-DA model, optimizing the number of latent variables (LV) by means of a Random Subsampling cross-validation. The Fisher's Ratio computed in the LV space was the figure of merit for selecting the best subset of features (i.e. sensors and temperatures). To prevent local optima and obtain a frequency of selection for every feature, GA were repeated until entropy saturation criteria was accomplished.

Finally, we built a calibration model considering the features which reached the significance threshold. This classifier was employed to predict the class of test samples. Additionally, a permutations test was performed to ascertain if obtained results were statistically significant.

Results

After feature selection performed by GA (see Figure 1), 7 conditions were included to train the model for CO identification. 86% Classification Rate, with 94% Sensibility and 85% Specificity, was obtained when evaluating test data. Additionally, the Area Under the ROC was 0.90. Figure 2 shows both calibration and external validation data in the space of the first two latent variables. After a permutation test of the labels, a p-value=0.007 was obtained.

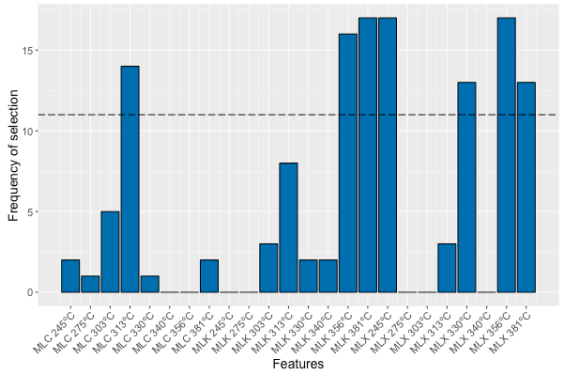

Fig. 1: Feature selection performed by GA, which yielded to the selection of 7 elements (313ºC for MLC, 356ºC, 381ºC for MLK, and 245ºC, 330ºC, 356ºC, 381ºC for MLX).

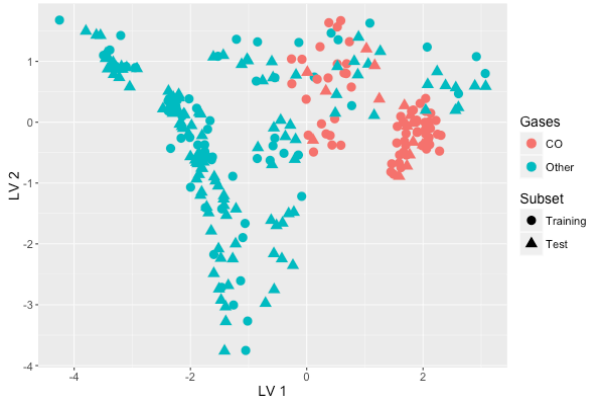

Fig. 2: Training and test data on the space of the first two latent variables of the final calibration model.

Conclusions

We showed that models trained with several sensor arrays also provide high performance when applied to new replicas of the sensor array. To do so, calibration needs to be performed over a number of replicas so the trained model is able to reject sensor variability. Moreover, we proposed a methodology to reduce the number of sensing elements so that similar or even higher performances can be achieved with a simpler calibration model. In summary, our approach would reduce calibration costs in a scenario of mass production, facilitating thereby mass deployment of gas sensor arrays.

References

[1] L. Fernandez, et al. *Sensors and Actuators B.* 231, 276-284 (2016); doi: 10.1016/j.snb.2016.02.131
[2] J. W. Gardner, et al. *Sensors and Actuators B*. 106, 114-121 (2005); doi: 10.1016/j.snb.2004.05.043